DOCUMENT RESUME

ED 412 263 TM 027 566

| | |
|---|---|
| AUTHOR | Csete, Josephine M.; Jones, John; Kwan, Kam-por |
| TITLE | Not Contradictory but Different; Understanding the Results of an Evaluation Study Employing an Integrated Design. |
| PUB DATE | 1997-03-00 |
| NOTE | 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997). |
| PUB TYPE | Reports - Research (143) -- Speeches/Meeting Papers (150) |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | *College Faculty; College Students; Confidentiality; Critical Theory; Evaluation Methods; Focus Groups; Foreign Countries; Higher Education; *Integrated Activities; Interviews; Qualitative Research; *Research Design; *Student Evaluation of Teacher Performance |
| IDENTIFIERS | *Hong Kong; Stakeholders |

ABSTRACT

This paper describes an evaluation study conducted on a student evaluation of teaching form and evaluation process newly instituted at a university in Hong Kong. The study used an integrated design with both quantitative and qualitative methods and collected information from a variety of stakeholders. The major focus of the paper is on the diverse data collected from one stakeholder group, the academic staff, (through focus group interviews in seven departments, individual interviews, and survey questionnaires completed by 212 faculty members) and the decisions and processes involved in reconciling and reporting what often appeared to be conflicting data. A framework including positivist (or quantitative), interpretive (qualitative), and critical theory approaches is used to explain the decisions made and process used. Two appendixes illustrate the reporting format for a single area of interest and the contributions of diverse data sources to the issue of confidentiality. (Contains two tables and six references.) (Author/SLD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED 412 263

# Not contradictory but different: Understanding the results of an evaluation study employing an integrated design

Josephine M. Csete

John Jones

Kam-por Kwan

Educational Development Unit
The Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong

Tel: (852) 2766-6317
E-mail: etjcsete@polyu.edu.hk

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Josephine Csete

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

## ABSTRACT

This paper describes an evaluation study conducted on a student evaluation of teaching form and process newly instituted at a university in Hong Kong. The study used an integrated design employing both quantitative and qualitative methods and collected information from a variety of stakeholders. The major focus of the paper is on the diverse data collected from one stakeholder group, academic staff, (through focus group and individual interviews and survey questionnaires) and the decisions and processes involved in reconciling and reporting what often appeared to be conflicting data. A framework including positivist (or quantitative), interpretive (qualitative) and critical theory approaches is used to explain the decisions made and process used.

A paper prepared for the American Educational Research Association Annual Meeting
Chicago, Illinois
March 26, 1997
(Paper Presentation No. 24.17 10:35 - 12:05 Sheraton Ballroom I, Level 4)

A copy of this paper can be downloaded from the Educational Development Unit Homepage at:
http://hednet.polyu.edu.hk/confpapers.html

TM027566



BEST COPY AVAILABLE

## Background

Hong Kong Polytechnic has been in existence since 1972, first granting sub-degree level diplomas and certificates, and later adding Bachelors and postgraduate degrees, in practical professional areas such as business, accounting, nursing, physiotherapy and engineering specialties. In 1994 it was granted university status and officially renamed "The Hong Kong Polytechnic University" (PolyU). It is the largest of the seven universities in Hong Kong with over 1,000 full-time faculty and approximately 20,000 (13,000 full-time equivalent) students.

The change to university status necessitated expansion into new areas and created new requirements. One of the changes decided upon in 1994 by the senior management at the university was that a system of individual staff appraisal for all academic and administrative staff was to be instituted in the 1995/96 academic year. In a related decision, senior management declared that a standardized instrument and set of procedures for gathering student feedback on teaching was also to be implemented and used in the appraisal process.

A working party was established to design the form and accompanying procedures, and developed an optical mark read questionnaire called the "Student Feedback Questionnaire" (SFQ). This consisted of: 1) three open-ended questions asking students to comment on "best," "worst," and "areas for improving" an individual staff member's teaching; 2) 18 closed-response questions divided into six areas to which students indicated their level of agreement on a five-point scale; and 3) a section for optional closed response questions added by the staff member or department.

The working group established requirements such as that all teaching staff were to use the SFQ to gather feedback from students in two classes each academic year. Similarly, guidelines were issued as to how the forms were to be administered and processed (with the intention of maintaining confidentiality and consistency across 26 academic departments). The working party also determined that a "pilot study" should be carried out that examined the usefulness of the SFQ form and administrative procedures so that improvements could be made.

It is the "pilot study" of the newly introduced Student Feedback Questionnaire (SFQ) form and associated procedures that provided the impetus for this paper. The three authors were the evaluators responsible for designing the study, collecting and analyzing the data, and writing the final report summarizing the data and making recommendations.[1] The remainder of this paper describes the evaluation study which involved an integrated design employing the positivist (quantitative), interpretative (qualitative) and critical theory (action oriented) approaches, before focusing on a smaller portion of the study (feedback from academic staff) to explore issues related to how the approaches were combined. Special attention will be given to illustrating the differences between the data sets, describing the processes used to help reconcile these differences, and the development of an appropriate reporting format.

---

[1] Refer to "Report on the Evaluation of the Implementation of the Student Feedback Questionnaire (SFQ)" by J. Csete, J. Jones and K. P. Kwan, October 1995. Published by the Educational Development Unit, The Hong Kong Polytechnic University.

## Study Design and Methods

Approaches contributing to the integrated design

Before designing the evaluation study it was essential to make the underlying assumptions of the evaluators and stakeholders explicit and identify resources and constraints. The three evaluators charged with designing and conducting the study held with Candy (1989) that it is both possible and, in this particular case, beneficial to design a study incorporating a wide range of approaches across the positivistic, interpretive and critical paradigms. The positivist paradigm would contribute quantitative methods which allow for collection of data from a larger number of stakeholders and expression of results as significant trends (which decision makers find helpful when determining policy). The interpretive tradition would contribute qualitative methods which include thick description detailing the complexity of the situation; explore stakeholders' motives, concerns and ideas which would not be captured in a quantitative study with predetermined variables of interest (but could suggest beneficial policy changes); and incorporate a commitment to involving the stakeholders in the study as participants. Rather than suggesting specific data collection and analysis methods, critical theory contributed an ever present awareness of the context; the potential ramifications of the evaluation study results; an acknowledgment of the diverse "interests" of the various stakeholder groups; and the resulting perspective that, whether or not desired, the evaluation study was an agent for change rather than an "objective piece of research." The critical paradigm therefore influenced study design and decisions throughout the data collection and analysis process.

The evaluators felt that the critical paradigm was inescapable in that the potential ramifications of the new student feedback form were large. The form was to be completed hundreds of thousands of times each year by students, thousands of hours of staff time would be required to administer and process the questionnaire, and the results could affect decisions to retain or promote more than one thousand teaching staff. Incorporating the critical approach would strengthen the evaluation study through acknowledging its political and individual implications. Candy (1989) quotes Popkewitz to illustrate how the critical paradigm goes beyond the interpretive in this respect:

> Whereas interpretive approaches may be inclined towards revealing misconceptions and confusion, while leaving situations unchanged, 'the function of critical theory is to understand the relations among value, interest, and action and, to paraphrase Marx, to change the world, not to describe it' (Popkewitz, 1984: 5).

The importance of the evaluation study also justified the choice of incorporating both quantitative and qualitative approaches into the integrated design. "Triangulation" -- the use of different methods to contribute different information and compensate for inherent weaknesses or bias of any single method -- would strengthen the study (Jick, 1979; Matheson, 1988). In addition, an integrated design was likely to yield a more complete picture of the complex situation to be studied (Csete and Albrecht, 1994; Creswell, Goodchild and Turner, 1996).

An integrated design was also feasible given the time, staff, and funding made available for the project. The evaluation team had experience in both quantitative and qualitative methods as well as knowledge of the three paradigms. Approximately six months were available in which to conduct the project. In addition to US$7500 used to pay for several months of research assistant help (mainly for conducting focus group and individual interviews) and tape transcription, the university provided clerical assistance and a portion of the three evaluators' time. The remaining major expense related to printing the surveys and final report and was absorbed by the university and unit conducting the evaluation.

## Study design

The evaluation study focused on four domains and collected data from six sources (see Table I). The domains were the major areas to be examined -- such as the form itself, and how the form was administered and the results used. The sources included stakeholder groups involved in the data collection, analysis and dissemination process as well as those who were being directly evaluated by the forms -- teaching staff. The study design could be characterized as "simultaneous triangulation" (Morse, 1991) as qualitative and quantitative methods were used concurrently with little interaction during the data collection period. The size of the evaluation study makes it difficult to succinctly report on the entire study in a meaningful way. Instead, this paper focuses upon methods used and data gathered on academic staff's responses to the three domains they were consulted on, as indicated in the shaded portion of Table I. This portion of the evaluation study is particularly appropriate as it employed the widest variety of data gathering techniques and provided the greatest challenges in reconciling seemingly disparate data.

**Table I:**
**Domains, sources and methods of the evaluation**

| *Sources* | *Domains* | | | |
|---|---|---|---|---|
| | **Administration guidelines and procedures** | **The questionnaire instrument** | **Reporting & use of the results** | **Inputting & processing of data** |
| **Dept. administrative staff** | Focus group interviews of 7 pilot departments | | Focus group interviews of 7 pilot departments | |
| **Academic staff** | **Focus group interview with staff from one pilot department<br>Survey questionnaire to all staff members in 7 pilot departments<br>Individual interviews with staff from 6 pilot departments** | | | |
| **Students** | Focus group interviews of 3 pilot departments | | | |
| **Empirical student feedback data** | | Statistical analysis of 1200+ class sets | | |
| **Educational support staff** | | | | Document analysis |
| **Computing support staff** | | | | Telephone interview |

## Data collection methods from academic staff

Three kinds of data were collected from academic staff. These were quantitative data from the "ticked boxes" of a survey questionnaire, qualitative data from the written comments on open-ended questions on the same survey, and qualitative data provided by the records (summary reports and transcripts) from individual and focus group interviews. Each of the three data collection methods are further described in Table II.

**Table II:**
**Methods of collecting data from academic staff**

1. **A focus group interview** was held with staff from one of the seven designated pilot departments (which had been nominated by the Deans as representative of the six Faculties on campus). The interview was semi-structured and focused on academic staff members' perceptions of the questionnaire items, procedures for administering the SFQ, and reporting and use of the results. This interview was useful in the design of the survey questionnaire described in item 2 below.

2. **A survey questionnaire** was administered to all academic staff in the seven pilot departments. The pilot group rather than all teaching staff on campus was surveyed, as surveying a smaller number both allowed for collection of more detailed data and greater follow up with respondents to ensure as good a response rate as possible. The questionnaire was distributed to 286 academic staff members, and after two follow-up procedures, 212 completed questionnaires were received (a response rate of approximately 74%).

The questionnaire was divided into five sections to gather information in the following areas:

a) General information on the respondent (5 closed items)
b) Guidelines for administering the SFQ (9 closed items)
c) The SFQ format, and the individual items (30 closed items)
d) The reporting and use of the SFQ results (7 closed items)
e) General comments (one open-ended question)

In addition to the closed items, there was room for respondents to write open-ended comments in each section. The final survey format consisted of a two large sheets folded and stapled to form eight 8 x 11" pages (including a one-page explanatory cover letter and another for return mailing).

3. **Individual interviews** were conducted with 25 staff members from across six pilot departments.[2] These staff members were nominated by their respective heads of department. Each interview (which typically lasted about an hour) was tape-recorded, transcribed, and translated into English in instances where the staff member wished to be interviewed in Chinese (Cantonese).

## The Journey Toward "Results"

### Disparate results from different data collection methods

As the form and its intended use in the new appraisal process had largely been constructed and implemented without consulting the teaching staff, at the start of the study the evaluators expected that the respondents would be critical and perhaps even hostile about the form and procedures. However, as the analysis process began with each evaluator focusing on only one of the data sources from this set, in the first team meeting in which data was reported they were surprised by the great disparity between the data sets.

The quantitative data from the survey suggested that, for every question asked, the staff were satisfied with the current situation! For each of the 37 closed questions asking an opinion, at least a majority of the respondents indicated they were satisfied with the system. Much more often, 70% or over chose answers indicating the system did not need changing. These results are particularly striking as for most questions three options were offered in which only one option ("Keep as it is") could be construed as

[2] As five members of the seventh pilot department had been involved in the focus group interview, a decision was made not to also conduct individual interviews in this department.

"positive" whereas the remaining two options ("Changes required" and "To be discarded") were regarded as expressions of dissatisfaction with the current form or process. If the quantitative data had been the sole source, the evaluators would have concluded that staff were generally satisfied with all questions on the new form itself as well as with everything related to how the form was administered and used in appraisal.

On the other hand, the qualitative data suggested a very different story. The written comments to the open-ended questions in the survey were overwhelmingly dissenting in content. Although it is not unusual for the larger proportion of optional written comments on questionnaires to represent dissenting viewpoints, the fact that each open-ended question (on a rather long questionnaire) elicited written comments from at least one third of the 212 respondents is indicative of how widespread the concerns were in this particular instance. Also, the depth of concern was exemplified by the length, (sometimes emotion laden) word choice, as well as sheer number of written comments. Even though the survey was intentionally designed to encourage respondents to provide written comments for any boxes ticked "changes required", the length, detail and number of written comments was surprising. One open-ended question elicited 282 individual comments. Although sometimes as much as half a page had been allotted for written comments, some respondents' comments overflowed into the margins or were submitted on additional sheets of paper. The data in the transcribed interviews and focus group report was also predominantly critical. There seemed to be greater disagreement, or at least expression of concern, the more "qualitative" the data collection method.[3] Interviews were more critical and went into greater detail than had occurred in the written comments portion of the questionnaires.

The disparity of information presented from different data sources is illustrated in Appendices A & B. The data results from the quantitative portion of the survey, written comments and focus group and individual interviews are presented in turn, followed by implications and suggestions (Appendices A & B are taken verbatim from the final report as examples of the reporting format on individual items). The purpose of the paper is not so much to focus on the specific issues depicted in the appendices, as to refer to these examples to illustrate how the types of data differed, and how these seemingly conflicting perspectives ultimately contributed to a fuller understanding of a complex situation.

The question as to whether it is appropriate for the SFQ to focus on collecting feedback on individual staff members rather than the subject or entire (three-year) course of study exemplifies both the seemingly contradictory results, and how each data set contributed different sorts of information that led to recommendations for changes that were more likely to improve the form, refine how it would be used in the future, and make life easier for the numerous people who were to fill out, process and make use of the SFQ (see Appendix A).

A closer examination of the issue highlights both the conflicts between data sets and the opportunities for fuller understanding (and it is hoped, better recommendations) that grew out of this conflict. The quantitative results show that a majority supported the status quo. This ratio of 61.3% of respondents ticking "Keep it as it is" to 38.7% ticking "Changes required" represents one of the very lowest "positive" scores. The open-ended comments demonstrate a very different view and express respondents'

---

[3] One possible explanation for this disparity between the quantitative and qualitative data sets is passive resistance. It occurred to the evaluators that staff members may not be accepting the new system, but due to concern about negative consequences, also not openly rejecting it. Rather, they intended to work around the system. In such a situation qualitative methods would actually bring into the open the negative attitudes that could remain masked in the quantitative portions of the survey. This theory would also account for the feedback getting progressively more negative, the more opportunity there was for providing details and following up on comments.

concerns.[4] The analysis of the written comments allows for the presentation of two broad areas of concern related to this issue, followed by a few actual written comments to get across the spirit of these concerns in the respondents' own words and, perhaps, remind the reader of the impact of this concern upon the individual. Qualitative interviews and focus groups with teaching staff and students (who were also consulted on this issue) support the written comments and play a role in delineating these concerns and demonstrating how widely they are held. (In all cases data from teaching staff interviews and focus groups supported the issues presented in the written comments.)

Another kind of benefit derived from exploring a single issue through diverse data collection methods is illustrated in Appendix B. This question on confidentiality is not an example of conflicting data. The ratio of 93.7% ticking "Keep it as it is" to 6.3% ticking "Changes required" is one of the highest quantitative measures indicating support, and the corresponding small number of written comments (16) suggests respondents felt less compelled to expound on this issue. The nature of the open-ended comments was more that of clarifying or suggesting enhancement rather than expressing dissension. An important discovery was that the written comments, again supported by staff interviews and refined by qualitative data collected from two other stakeholder groups, gave useful information which could not possibly have been gleaned from the quantitative portion of the study.

Data collected through qualitative methods revealed respondents had different concepts of "confidentiality" as well as different references as to whose confidentiality was to be protected. Although staff generally agreed that it was important to maintain confidentiality, many felt it would be personally useful, as well as useful for departments, to have some measure of where they stood by comparison to consolidated scores. Teaching staff also indicated they felt sensitivity to the context of the teaching ("size of class", "year of study" are examples) was so important it overrode the original principle of maintaining "strict confidentiality" which prevents the inclusion of any contextual factors that could potentially lead to identification of individual teachers or departments.

Another issue raised by the qualitative data was whether student or staff member confidentiality was being referred to. Asking the question about confidentiality in interviews and focus groups allowed different stakeholder groups to bring up concerns related to also protecting students' confidentiality. The issues that arose from the qualitative portion of the evaluation study didn't really conflict with the quantitative data. Rather, approaching a variety of stakeholders with an open-ended question on this issue allowed previously unthought of concerns to be voiced, and ultimately, changes made to the form and system that streamlined the process and was likely to make results more useful.

<u>The process of reconciling the data</u>

> The presentation of data is one thing, the drawing of conclusions and recommendations from the multiple data sources is another. It is necessary to outline the procedures that were used to arrive at conclusions and recommendations, and to make explicit the value positions that underpin the conclusions that have been drawn.
>
> [Final Report, p.10]

---

[4] Not surprisingly, the number of written comments (70) is relatively more than for most single items. This corresponds with the greater quantitative degree of dissension.

The idea that such different types of data could be coalesced without compromising the spirit of any single source was far from apparent the first time data results were shared. As mentioned earlier, data collection and early analysis had gone on largely independently with each of the three evaluators concentrating on a different data set. Imagine the surprise of both when the evaluator who had been analyzing written comments said they were overwhelmingly very critical and the evaluator analyzing quantitative questions responded that every single one of the quantitative items indicated a majority supported all aspects of the form and process. (The evaluator who had been examining written comments went so far as to question whether the quantitative measures had been accidentally coded in reverse!) Although the evaluators were familiar with the literature on the reasons for and benefits from integrated studies and an integrated design had made intuitive sense, they were initially at a loss as to how such seemingly contradictory forms of data could be reconciled to arrive at a balanced representation that did not compromise the integrity of any of the data sets.

The literature does not yet provide much detailed advice as to how the data in integrated studies are to be "integrated." What follows is a brief description of the evaluation team's steps and process for integrating the data. Often the team was figuring out what to do as it progressed. However, a set of principles guided these decisions. Many of these principles were drawn from critical theory, though intuitively via value positions, rather than as a set of algorithmic prescriptions. The following quote from the final report illustrates how this paradigm contributed to important decisions during the analysis process:

> "It is important, we feel, that *as many staff as possible*, feel *as comfortable as possible* with the procedures that are being used to appraise their teaching.
>
> This leads us to look beyond a simple majority vote, and to seek modifications and fine tuning that:
>
> - preserve the essential substance of the status quo (with which most staff are satisfied), and
> - would address as many as possible of the concerns expressed by the minority of staff, concerning the existent procedures.
>
> With this in mind we then looked at the qualitative data included in the open-ended comments, the individual interviews and the various focus group meetings, to search for the 'optimal' modification..."
> [emphases in original] [Final Report, p. 11]

The first and largest component of the data analysis process involved a series of meetings between the three evaluators. The design of the entire evaluation study (see Table I) proved a useful organizing principle. Each meeting focused on a sub-portion of one of the four domains (administration, instrument, use of results, data processing). Evaluation team members divided up responsibility for becoming familiar with one or more of the data sets and arriving at results on that single sub-portion prior to the designated meeting. This meant that each meeting focused on data from all stakeholders consulted and collected through all methods, but for a relatively narrowly defined issue. Adequate time was scheduled for each meeting (often half a day). The following procedure quickly evolved for meetings:

1) Each meeting started with reviewing findings from the previous meeting (often with a presentation and critique of the first draft of recommendations).

2) Each of the three evaluators presented their results, with some illustrative substantiation, for the data set(s) and stakeholders they had examined for the new issue.
3) Lengthy discussions were held in which areas of "easy" agreement were first decided and put aside before addressing points of seeming conflict one by one.
4) After several hours of concentrated effort the evaluators usually found a trend or conceptual framework that accounted for most of the conflicts in the data on the single issue being examined and tested this trend for applicability across the issue.
5) Near the end of most meetings there was a brief discussion on how what had been discovered in this sub-portion suggested improvements to the reporting format (which was simultaneously evolving).
6) Meetings ended with agreement on who was to write up the first draft on the present sub-portion, the next sub-portion to be examined and assignment of data sets for analysis. (Evaluators tended to stick with particular data sets across all meetings).

After the evaluation team had completed analysis of each of the sub-portions and further revised initial rough drafts to come up with a comprehensive report that they felt linked the issues and was reported in a format that appropriately represented the disparate data sets (see below for details), they presented it to the professional staff of their unit. Taken together the ten staff represented a wide range of specializations in education. An added benefit was their familiarity with the university and stakeholder groups. As seven of these staff had not been involved in the analysis process, they were also more likely to catch inconsistencies and suggest ways to further clarify the presentation.

Arriving at an appropriate reporting format

Just as the evaluation team had to figure out the process for data analysis in this integrated study, they also had to develop a format appropriate for reporting the results and recommendations. It was important to come up with a format that represented what had been learned from disparate data collection methods and stakeholder groups without emphasizing one at the expense of another. It was also challenging to arrive at a format that presented a complex situation from multiple perspectives without confusing the reader. As alluded to in the section above, the reporting format was constantly being revised and refined over the course of data analysis. Each of the three approaches that had contributed to study design and data analysis decisions also influenced the reporting format. Critical theory will be considered first.

Certain "givens" colored the evaluation study. It was clear at the commencement of the study that some form of student evaluation of teaching would be used at the university, that a standardized form would be part of it, and that results of this instrument would be used in a staff appraisal process which would be both developmental and judgmental. Critical theory suggested that these givens be taken into account to arrive at recommendations that were most likely to effect positive change.[5] A corollary of this was that the results needed to be presented in a way that "made sense" to the decision makers. The scientific and engineering backgrounds of the decision makers (and their consequential preference for quantitative data) was therefore taken into account when designing the report format. For each of the individual items, quantitative data were reported first, but quickly followed by summaries of trends evident from qualitative sources (with the number of comments always included). A few qualitative quotes

[5] This is very different from "accepting" these givens and avoiding challenging them in the study. There were several instances in which the data suggested changes challenging these givens, and these changes were also suggested in the final report.

helped personalize the issues. Summaries from all remaining stakeholders and data collection methods on the particular topic indicated that the concerns were widespread and could not be dismissed as particular to a single data collection method or "a few pranksters" (refer to Appendices I & II). Before presenting any of the results the report candidly acknowledged the disparity between quantitative and qualitative sources and listed the value positions used to interpret the data, including that of proposing "optimal modifications" even though the majority had not dissented (as quoted earlier in this paper).

Both the positivist and interpretive traditions contributed ideas as to how to present the different forms of data, as well as respect for each form of data while taking the weaknesses associated with the method into account. The standard of controlling for bias in the positivist paradigm also constantly challenged the evaluators to acknowledge and control their own convictions of what constituted an ethical and educationally sound instrument and process for assessing student feedback on teaching.[6] Advice from educational experts was labeled as such.

The final report was 34 pages plus appendices and contained the following sections:

1) Executive summary of recommendations for changes which concludes with a half page of "general issues" highlighted by the evaluation study, but left unresolved (three pages). Each recommendation was referenced to one or more pages in the body of the report.
2) Introduction citing official memoranda to describe how and why the evaluation study had been commissioned (two pages).
3) Description of study framework (including a version of Table I) and methods employed in the study (three pages).
4) Presentation of value positions and associated procedures used to draw implications (one page).
5) Examination of each issue addressed in the study, grouped according to study framework. Data from each of the methods used and stakeholder groups consulted for a given issue was presented followed by "implications and suggestions" (format displayed in Appendices I & II) (24 pages).
6) Five appendices containing the original SFQ instrument and copies of all data collection instruments used in the evaluation study.

## Lessons Learned

If the evaluation study were to be assessed on the basis of whether the results were judged credible and recommendations were ratified by decision makers it would certainly be a success -- all but two of the 17 recommendations were accepted.[7] As the evaluators, we also considered the study a very beneficial experience -- we had never before had to wrangle with such a large amount of "unruly" data that refused to fall neatly into place. When faced with questions that seem appropriate to an integrated design we will certainly do all we can to arrange for adequate time, money and staffing to carry out an integrated study.

Of course, there are some things we would readily repeat and others we would do differently based on what we learned from this experience. First, we will continue to make use of pilot groups. By focusing on a smaller group, and explaining that they

---

[6] For example, a significant portion of teaching staff expressed the opinion that students were not qualified and/or mature enough to assess their teaching. This trend was reported even though the evaluators disagreed with it.

[7] Not surprisingly, one of the rejected recommendations involved a financial commitment to create a university position to help administer the system.

were charged with "speaking for the many," pilot group respondents seemed much more willing to participate. An example is the 74% response rate for a survey in an institution where other surveys distributed to all staff have rarely managed to break into the teens in percentage of returns. An added benefit of focusing on a smaller pilot group was the ability to collect more detailed (and qualitative) data from them.

Second, we will make greater use of focus groups early in the study. Conducting the focus group with teaching staff prior to issuing the survey was very helpful. Ideas gleaned from the focus group resulted in many revisions to the questionnaire including adding questions in areas previously not considered, making the response format easier to follow, and clarifying the importance of the pilot group's participation.

Third, if repeating a similar type of study we would make less use of individual interviews. Individual interviews were not well suited to the purposes of this evaluation study, which needed to identify broad trends and arrive at recommendations for action, rather than unfold the complexity of staff views on teaching evaluation forms and processes. The interviews were time intensive to conduct, and especially to translate and transcribe.[8] It is beneficial to do a few interviews, but the advantages derived did not justify the effort expended for 25. What was learned from the interviews added nothing new to what had been gathered (much more easily) from the other qualitative data collection methods of written comments and focus groups. Interviews did differ in the level of detail they provided and the frequently greater expression of dissension. In a similar situation, conducting more focus groups would yield a similar amount of information with less effort. We would however, wish to do some interviews to check whether findings carry across data collection methods and to take advantage of the opportunity to collect greater detail and follow up on ideas.

We will continue to have more than two people involved in designing and conducting integrated studies. The study reported in this paper just happened to have three evaluators, but this turned out to be very fortunate. In situations such as this in which different data sets present such contrary positions, it is valuable (probably even essential) to have "champions" representing each perspective. However, the hours of discussion required to bring these disparate perspectives to consensus would have been much more difficult if two evaluators had found themselves pitted against each other. The interactions of three evaluators prevented this.

The study also served as a reminder to closely monitor our subjectivity and role as stakeholders in the issues under examination. Much has been written about the pros and cons of having evaluators from inside versus outside the system. The available resources precluded employing an external evaluator. As insiders we had the advantage of already knowing about the system to be evaluated and the culture of the university. And as insiders we also had the disadvantage of having to identify and control for our own interests. Although we had not developed the SFQ form or the procedures associated with it, we were to be given the responsibility for administering it university-wide. In the future we hope to include a few "critical friends" from outside the unit responsible for conducting the study to act as consultants in helping us identify and control subjectivity.

Finally and perhaps most importantly, conducting this evaluation study convinced us that an integrated approach was much more likely to yield an accurate picture of phenomena of interest. It is ironic that based on what was learned from both quantitative and qualitative data, we were left wondering about the accuracy and usefulness of the largely quantitative form that had been the focus of the evaluation

---

[8] Interviews may have been especially problematic in the given context. The majority of staff chose to be interviewed in Chinese, requiring additional effort and expense to translate the interviews.

study.[9] We were concerned that judgments were to be made about a teacher's effectiveness, based on a single perspective (students) and a single data collection method. This experience in the benefits of an integrated evaluation study has spurred us to propose an integrated approach to the assessment of teaching itself. A current project is the development and piloting of a teacher's portfolio -- in which the SFQ is only one of many types of data collected from multiple sources to build an accurate picture of an individual's teaching performance and make recommendations for further improvement.

## References


Candy, P. C. (1989). Alternative paradigms in educational research. *Australian Educational Researcher* 16(3): 1-11.

Creswell, J. W., Goodchild, L. F. & Turner, P. P. (1996). Integrated qualitative and quantitative research: Epistemology, history, and designs. *Higher education: Handbook of theory and research* (Vol. XI) New York: Agathon Press.

Csete, J. M., & Albrecht, R. R. (1994, December). *The best of both worlds: Synthesizing quantitative and qualitative research in the primary care setting.* Paper presented to the Primary Care Research Methods and Statistics Conference, San Antonio, Texas.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly* 24: 602-611.

Mathison, S. (1988). Why triangulate*?* *Educational Researcher* 17(2): 13-17.

Morse, J. M. (19[illegible]). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research* 40(1): 120-123.


[9] This is not a problem particular to our University. The form evaluated in this study is similar to those used to collect student feedback on teaching at universities around the world.

## Appendix A:
## Full reporting format for a single area of interest

**Q1 The SFQ focuses on collecting feedback on individual staff members rather than the subject or course.**

4.1
- 61.3% of respondents ticked "Keep as it is".
- 38.7% ticked "Changes required".

4.2 This item attracted the largest number of open-ended comments in this Section of the staff survey; 70 in all. Most of these can be grouped into two broad categories as follows.

(a) There is a need to focus on course, subject, teaching teams, and the overall structure of the programme of study, in addition to the individual teacher. 'Many departments and staff have been carrying out formative evaluations, designed to monitor the quality of courses and subjects for some time. This has been very valuable. In many ways the SFQ, with its judgmental focus on the individual teacher which is designed to inform the staff appraisal system, subverts this.

(b) Teachers do not have total control of syllabuses, or course and subject arrangements, though student reactions to these matters will inevitably colour their judgments of the quality of the teaching. Student feedback has to be interpreted in context; and part of that context (in addition to the factors already mentioned) includes class size, the level of the course, etc.

Actual written comments that capture the essence of these two points are as follows.

*"It is more important to focus on the quality of the subject or the course rather than individual staff members if it is for developmental purposes. Indication of individual staff members performances can be incorporated into a questionnaire based on evaluation of a subject and course."*

*"We need to know about course content, design and balance, about the assignments and their content and timing. We can change these things easily. Personality and interactive style are harder to change."*

*"Background information on the student, e.g. attendance, academic records etc. are to be collected to assess the validity of a student's answer. Environmental factors, e.g. students' comments on course syllabus, classroom, time-tabling, etc. should be considered as well."*

*"The SFQ should aim to collect student feedback on both individual staff and the nature of the subject or course, and even the planning/timing of the subject, because sometimes it could be the problem of administration policy or the planning of the programme."*

4.3 Interviews with academic staff – both individual and in focus group – reflected these opinions, and elaborated them along the following lines.

For formative evaluation for developmental purposes, "imperfections" in questionnaire design and administration procedures that result in summaries of results that are not "completely accurate" are acceptable. The information is contextual and is intended to be used in that context to make changes for the better. However, for summative judgmental purposes, these context factors are crucial, and have to be taken into account in order to interpret the student feedback appropriately.

4.4 Students, in commenting on this kind of issue, stressed how confusing it was to have to complete three different kinds of questionnaire forms, often in the same class: the SFQ, a course evaluation and a subject evaluation. In addition, students reported having to fill in as many as ten or more separate SFQ forms. (The record number brought to our attention was 14!)

## Appendix A (cont'd):
## Full reporting format for a single area of interest

4.5 Implications and suggestions

(a) SFQ and course/subject evaluations have to be co-ordinated if both the developmental and judgmental aspects are to be maintained. It is practically impossible to prescribe detailed arrangements that will fit all circumstances. This is something for departments to consider and co-ordinate.

(b) Context factors have to be included as part of the overall evaluation, in order to enable the feedback on the individual staff member to be interpreted appropriately. The suggestion is that a "cover sheet" should be attached to each class set of student feedback data, so that information relating to class size, teaching mode, level of the course, etc. can be included in the interpretation of the data.

[There are some implications here for confidentiality — see Q6; however it should be easy to build in safeguards that are consistent with those common in other institutions around the world.]

## Appendix B:
## Contributions of diverse data sources to the issue of confidentiality

### Q6 Confidentiality in collecting, processing and reporting of data is necessary.

4.22
- 93.7% of respondents ticked "Keep as it is".
- 6.3% ticked "Changes required".

4.23 There were 16 open-ended comments; a variety of issues was raised, including whether confidentiality was principally concerned with staff or students and also the need to publish aggregated data. For example:

> *"Confidentiality of students must be maintained. However a system of identification should be implemented to make students accountable for their responses, to Head of Department or senior management."*
>
> *"The best scores should be announced."*
>
> *"Basically keep it as it is, but also publish a departmental profile of scores and a PolyU-wide profile. This could be a useful benchmark. Also what about publishing the highest and lowest scores PolyU-wide for each question (class summation divided by the no of students)."*

4.24 Although there were not many open-ended responses on the questionnaire forms, a number of pertinent issues emerged during the interviews.

(a) If the feedback is for summative, judgmental purpose then it should definitely be confidential — and in fact some respondents suggested that it should go to the staff member first (before being seen by the Head of Department) in this case. On the other hand, if the feedback is primarily for formative purposes, then confidentiality is not such an issue.

(b) There was a general consensus that while confidentiality is necessary with respect to individual feedback, this principle need not (indeed should not) apply to aggregated data. There is a general feeling that it would be useful to know, generally, what kinds of feedback scores other staff are getting.

This point relates to the discussion about "context factors" in Q1 above. It would be possible to compile average ratings for different kinds of teaching contexts; but then the very strict current confidentiality arrangements would need to be relaxed slightly, in order to enable "context" data to be included as part of the data set. The essential question is "How confidential does 'Confidential' need to be?"

4.25 A number of issues have been raised in the focus group interview for departmental administrative staff:

(a) The insistence on "absolute" confidentiality in distributing, collecting and reporting of the SFQ data has led to the use of arbitrary staff and class coding on the SFQ forms, which greatly complicated the administration, processing and reporting procedures. A substantial amount of manpower has been devoted to assigning codes, instructing students to put down those arbitrarily assigned codes, and checking that the codes have been entered accurately on the SFQ forms. It also opened up sources of potential errors in the implementation process which might have significantly affected the accuracy of the results. Furthermore, the use of arbitrarily assigned codes in reporting the SFQ data has made the results extremely obscure, and impossible to identify in case the departmental record of coding is lost or unavailable.

BEST COPY AVAILABLE

## Appendix B (cont'd):
## Contributions of diverse data sources to the issue of confidentiality

(b) Administrative staff also raised the matter of some teaching staff staying in the room while the questionnaire was administered. Some students seemed to feel uneasy about this. More generally it became apparent that there was not much consistency in the way in which the forms were administered — despite very detailed guidelines being issued.

4.26 Students were of the opinion that the staff member concerned should be absent from the room when the SFQ forms are distributed and collected. Some students hoped that results should be made public — at least at an aggregated departmental level — and some indication given as to what action was being taken on the basis of the feedback that they provided.

4.27 Implications and suggestions

(a) Basically, confidentiality in collecting, processing and reporting of the data should be maintained. However, it makes more sense to use actual staff name/code and actual subject code on the SFQ forms and in reporting the SFQ results. This greatly reduces the workload in a administering the SFQ forms, minimises the chance of human errors in handling the forms, and makes the SFQ results more "visible".

(b) It is definitely undesirable to publish the SFQ result of the individual staff members. However, for students to treat the SFQ exercise seriously, they must convinced that their feedback is valued and may lead to changes or improvements. It is therefore imperative that students are briefed adequately on the objective of the SFQ and the possible use of the student feedback data before the SFQ exercise.

Departments should be encouraged to devise their own way of briefing their students.